Reed Muller Sensing Matrices and the LASSO

Robert Calderbank and Sina Jafarpour

Abstract

We construct two families of deterministic sensing matrices where the columns are obtained by exponentiating codewords in the quaternary Delsarte-Goethals code $DG(m,r)$. This method of construction results in sensing matrices with low coherence and spectral norm. The first family, which we call Delsarte-Goethals frames, are 2^m - dimensional tight frames with redundancy 2^{rm}. The second family, which we call Delsarte-Goethals sieves, are obtained by subsampling the column vectors in a Delsarte-Goethals frame. Different rows of a Delsarte-Goethals sieve may not be orthogonal, and we present an effective algorithm for identifying all pairs of non-orthogonal rows. The pairs turn out to be duplicate measurements and eliminating them leads to a tight frame. Experimental results suggest that all $DG(m,r)$ sieves with $m \leq 15$ and $r \geq 2$ are tight-frames; there are no duplicate rows. For both families of sensing matrices, we measure accuracy of reconstruction (statistical $0-1$ loss) and complexity (average reconstruction time) as a function of the sparsity level k. Our results show that DG frames and sieves outperform random Gaussian matrices in terms of noiseless and noisy signal recovery using the LASSO.

Index Terms

Compressed Sensing, Reed-Muller Codes, Delsarte-Goethals Set, Random Sub-dictionary, LASSO

I. Introduction

The central goal of compressed sensing is to capture attributes of a signal using very few measurements. In most work to date, this broader objective is exemplified by the important special case in which the measurement data constitute a vector $f = \Phi\alpha + e$, where Φ is an $N \times \mathcal{C}$ matrix called the *sensing matrix*, α is a signal in $\mathbb{C}^{\mathcal{C}}$, that is well-approximated by a k-sparse vector (a signal with at most k non-zero entries), and e is additive measurement noise.

The role of random measurement in compressive sensing (see [1] and [2]) can be viewed as analogous to the role of random coding in Shannon theory. Both provide worst-case performance guarantees in the context of an adversarial signal/error model. In the standard paradigm, the measurement matrix is required to act as a near isometry on all k-sparse signals (this is the Restricted Isometry Property or RIP introduced in [3]). It has been shown that if a sensing matrix satisfies the RIP property then Basis pursuit [1], [4] programs can be used to estimate the best k-term approximation of any signal in $\mathbb{C}^{\mathcal{C}}$, measured in the presence of any ℓ_2 norm bounded measurement noise [5].

It is known that certain probabilistic processes generate sensing matrices that for $k = O(N)$ satisfy k-RIP with high probability (see [6]). This is significantly different from the best known results for deterministic sensing matrices [7] where k-RIP is known only for $k = O(\sqrt{N})$. We normalize the columns of a sensing matrix to have unit ℓ_2 - norm and define the worst case coherence μ to be the maximum absolute value of an inner product of distinct columns. It follows from the Welch bound [8] that $\mu \geq O\left(\frac{1}{\sqrt{N}}\right)$. When

Department of Electrical Engineering and Department of Mathematics. Princeton University. *calderbk@math.princeton.edu.*
Department of Computer Science. Princeton University. *sina@cs.princeton.edu*.

The work of R. Calderbank and S. Jafarpour is supported in part by NSF under grant DMS 0701226, by ONR under grant N00173-06-1-G006, and by AFOSR under grant FA9550-05-1-0443.

$\mu = O\left(\frac{1}{\sqrt{N}}\right)$ it then follows from the Gerschgorin Circle Theorem [9] that the sensing matrix satisfies k-RIP with $k = O\left(\mu^{-1}\right)$. In general however no polynomial-time algorithm is known for verifying that a sensing matrix with the worst-case coherence μ satisfies k-RIP with $k = \Omega\left(\mu^{-1}\right)$.

The RIP property is not an end in itself. It provides guarantees for a particular method of signal reconstruction, but there is significant interest in structured sensing matrices and alternative reconstruction algorithms. One example is the adjacency matrices of expander graphs [10], [11] where it is known to be impossible to satisfy RIP with respect to the ℓ_2 norm [12]. Sparse signal recovery is still possible with Basis Pursuit since the adjacency matrix acts like a near isometry on k-sparse signals with respect to the ℓ_1 norm. However error estimates are looser than corresponding estimates for random sensing matrices and resilience to measurement noise is limited to sparse noise vectors.

The coherence between rows of a sensing matrix is a measure of the new information provided by an additional measurement. The coherence between columns of a sensing matrix is fundamental to deriving performance guarantees for reconstruction algorithms such as Basis Puruit. There are two fundamental measures of coherence: The worst-case coherence μ which measures the maximal coherence between the columns of the sensing matrix, and the spectral norm $\|\Phi\|_2$ which measures the maximal coherence between the rows of the frame. The ideal case is when worst case coherence between columns matches the Welch bound $\left(\mu = O\left(\frac{1}{\sqrt{N}}\right)\right)$ and different measurements are orthogonal. Then, with high probability a k-sparse vector has a unique sparse representation [13], and this representation can be efficiently recovered using a LASSO program [14]. Section §II introduces notation and reviews prior work on the geometry of sensing matrices and the performance of the LASSO reconstruction algorithm.

In this paper we consider sensing matrices based on the $\mathbb{Z}_4$-linear representation of Delsarte Goethals codes. The columns are obtained by exponentiating codewords in the quaternary Delsarte-Goethals code; they are uniformly and very precisely distributed over the surface of an N-dimensional sphere. Coherence between columns reduces to properties of these algebraic codes. Section §II reviews the construction of Delsarte-Goethals (DG) sets of $\mathbb{Z}_4$-linear quadratic forms which is the starting point for the construction of the corresponding codes; each quadratic form determines a codeword where the entries are the values taken by quadratic form. Section §III introduces Delsarte-Goethals frames and Delsarte-Goethals sieves; the columns of these sensing matrices are obtained by exponentiating DG codewords. We then determine the worst case coherence and spectral norm for these sensing matrices.

Candès and Plan [14] specified coherence conditions under which a LASSO program will successfully recover a k-sparse signal when the k non-zero entries are above the noise variance. We use these results to provide an average case error analysis for stochastic noise in both the data and measurement domains. The Delsarte Goethals (DG) sensing matrices are essentially tight frames so that white noise in the data domain maps to white noise in the measurement domain.

Section §IV presents the results of numerical experiments that compare DG frames and sieves with random Gaussian matrices of the same size. The SpaRSA package [15] is used to implement the LASSO recovery algorithm in all cases. DG frames and sieves outperform random matrices in terms of probability of successful sparse recovery but reconstruction time for the DG sieve is greater than that for the other sensing matrices. We remark that there are alternative fast reconstruction algorithms that exploit the structure of DG sensing matrices. The witnessing algorithm proposed in [16] requires less storage, provides support-localized detection, and does not require independence among the support entries. On the other hand, LASSO reconstruction tends to be more robust to noise in the data domain.

II. Background and Notation

This Section introduces notation and reviews the theory of sparse reconstruction.

A. Notation

Given a vector $v = (v_1, \cdots, v_n)$ in $\mathbb{R}^n$, $\|v\|_2$ denotes the Euclidean norm of v, and $\|v\|_1$ denotes the ℓ_1 norm of v defined as $\|v\|_1 \doteq \sum_{i=1}^{n} |v_i|$. We further define $\|v\|_\infty \doteq \max\{|v_1|, \cdots, |v_n|\}$, and $\|v\|_{\min} \doteq \min\{|v_1|, \cdots, |v_n|\}$. Also the Hamming weight of v is defined as $\|v\|_0 \doteq \{i : v_i \neq 0\}$. Whenever clear from the context, we drop the subscript from the ℓ_2 norm. Also $v_{i\to j}$ denotes the vector v restricted to entries $i, i+1, \cdots, j$, that is $v_{i\to j} \doteq (v_i, v_{i+1}, \cdots, v_j)$.

Let A be a matrix with rank r. We denote the conjugate transpose of A by $A^\dagger$. Let $\boldsymbol{\sigma} = [\sigma_1, \cdots, \sigma_r]$ denote the vector of the singular values of A. The spectral norm $\|A\|$ of a matrix A is the largest singular value of A: that is $\|A\| \doteq \|\boldsymbol{\sigma}\|_\infty$. The condition number of Φ is the ratio between its largest and its smaller singular values: $\varsigma(A) \doteq \frac{\|\boldsymbol{\sigma}\|_\infty}{\|\boldsymbol{\sigma}\|_{\min}}$. Finally the nuclear norm of A, denoted as $\|A\|_1$ is the ℓ_1 norm of the singular value vector $\boldsymbol{\sigma}$.

Throughout this paper we shall use the notation φ_j for the j^{th} column of the sensing matrix Φ; its entries will be denoted by $\varphi_j(x)$, with the row label x varying from 0 to $N-1$. In other words, $\varphi_j(x)$ is the entry of Φ in row x and column j. We denote the set $\{1, \cdots .\mathcal{C}\}$ by $[\mathcal{C}]$. Let S be a subset of $[\mathcal{C}]$. Φ_S is obtained by restricting Φ to those columns that are listed in S.

A vector $\alpha \in \mathbb{R}^{\mathcal{C}}$ is k-sparse if it has at most k non-zero entries. The support of the k-sparse vector α, denoted by Supp(α), contains the indices of the non-zero entries of α. Let $\pi = \{\pi_1, \cdots, \pi_{\mathcal{C}}\}$ be a uniformly random permutation of $[\mathcal{C}]$. In this paper, our focus is on the average case analysis, and we always assume that α is a k-sparse signal with Supp$(\alpha) = \{\pi_1, \cdots, \pi_k\}$. We further assume that conditioned on the support, the values of the k non-zero entries of α are sampled from a distribution which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^k$.

B. Incoherent Tight Frames

An $N \times \mathcal{C}$ matrix Φ with normalized columns is called a dictionary. A dictionary is a tight-frame with redundancy $\frac{\mathcal{C}}{N}$ if for every vector $v \in \mathbb{R}^{\mathcal{C}}$, $\|\Phi v\|^2 = \frac{\mathcal{C}}{N}\|v\|^2$. If $\Phi\Phi^\dagger = \frac{\mathcal{C}}{N}\mathrm{I}_{N\times N}$, then Φ is a tight-frame with redundancy $\frac{\mathcal{C}}{N}$ (see [17]).

Proposition 1. *Let Φ be an $N \times \mathcal{C}$ dictionary. Then $\|\Phi\|^2 \geq \frac{\mathcal{C}}{N}$, and equality holds if and only if Φ is a tight frame with redundancy $\frac{\mathcal{C}}{N}$.*

Proof: Let Let $\boldsymbol{\sigma}$ be the singular value vector of Φ. We have

$$\|\Phi\|^2 = \|\boldsymbol{\sigma}\|_\infty^2 \geq \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 = \frac{1}{N}\mathrm{Tr}\left(\Phi\Phi^\dagger\right) = \frac{\mathcal{C}}{N}. \tag{1}$$

The inequality in Equation (1) changes to equality if and only if all the eigenvalues of $\Phi\Phi^\dagger$ are equal to $\frac{\mathcal{C}}{N}$. This is equivalent to the requirement $\Phi\Phi^\dagger = \frac{\mathcal{C}}{N}\mathrm{I}_{N\times N}$. ■

The mutual coherence between the columns of an $N \times \mathcal{C}$ sensing matrix is defined as

$$\mu \doteq \max_{i\neq j}\left|\varphi_i^\dagger\varphi_j\right|. \tag{2}$$

Strohmer and Heath [8] showed that the mutual coherence of any $N \times \mathcal{C}$ dictionary is at least $\frac{1}{\sqrt{N}}$. Designing dictionaries with small spectral norms (tight frames in the ideal case), and with small coherence $\left(\mu = O\left(\frac{1}{\sqrt{N}}\right)\right.$ in the ideal case$\left.\right)$ is useful in compressed sensing for the following reasons.

Uniqueness of Sparse Representation ($\boldsymbol{\ell_0}$ minimization) The following results are due to Tropp [13] and show that with overwhelming probability the ℓ_0 minimization program successfully recovers the original k-sparse signal.

Theorem 1. *Assume the dictionary Φ satisfies $\mu \leq \frac{c}{\log \mathcal{C}}$, where c is an absolute constant. Further assume $k \leq \frac{c\,\mathcal{C}}{\|\Phi\|^2 \log \mathcal{C}}$. Let S be a random subset of $[\mathcal{C}]$ of size k, and let Φ_S be the corresponding $N \times k$ submatrix. Then there exists an absolute constant c_0*

$$\Pr\left[\left\|\Phi_S^\dagger \Phi_S - I\right\| \geq c_0 \left(\mu \log \mathcal{C} + 2\sqrt{\frac{\|\Phi\|^2\, k}{\mathcal{C}}}\right)\right] \leq 2\,\mathcal{C}^{-1}.$$

Theorem 2. *Assume the dictionary Φ satisfies $\mu \leq \frac{c}{\log \mathcal{C}}$, where c is an absolute constant. Further assume $k \leq \frac{c\,\mathcal{C}}{\|\Phi\|^2 \log \mathcal{C}}$. Let α be a k-sparse vector, such that the support of the k nonzero coefficients of α is selected uniformly at random. Then with probability $1 - O\left(\mathcal{C}^{-1}\right)$ α is the unique k-sparse vector mapped to $u = \Phi\alpha$ by the measurement matrix Φ.*

Sparse Recovery via LASSO (ℓ_1 minimization) Uniqueness of sparse representation is of limited utility given that ℓ_0 minimization is computationally intractable. However, given modest restrictions on the class of sparse signals, Candès and Plan [14] have shown that with overwhelming probability the solution to the ℓ_0 minimization problem coincides with the solution to a convex lasso program.

Theorem 3. *Assume the dictionary Φ satisfies $\mu \leq \frac{c}{\log \mathcal{C}}$, where c is an absolute constant. Further assume $k \leq \frac{c_1\,\mathcal{C}}{\|\Phi\|^2 \log \mathcal{C}}$, where c_1 is a numeric constant. Let α be a k-sparse vector, such that*

1) *The support of the k nonzero coefficients of α is selected uniformly at random.*
2) *Conditional on the support, the signs of the nonzero entries of α are independent and equally likely to be -1 or 1.*

Let $u = \Phi\alpha + e$, where e contains N iid $\mathcal{N}(0, \sigma^2)$ Gaussian elements. Then if $\|\alpha\|_{\min} \geq 8\sigma\sqrt{2\log \mathcal{C}}$, with probability $1 - O(\mathcal{C}^{-1})$ the lasso estimate

$$\alpha^* \doteq \arg\min_{\alpha^+ \in \mathbb{R}^{\mathcal{C}}} \frac{1}{2}\|u - \Phi\alpha^+\|^2 + 2\sqrt{2\log \mathcal{C}}\,\sigma^2\,\|\alpha^+\|_1$$

has the same support and sign as α, and $\|\Phi\alpha - \Phi\alpha^*\|^2 \leq c_2\,k\,\sigma^2$, where c_2 is a numeric constant.

Stochastic noise in the data domain. The tight-frame property of the sensing matrix makes it possible to map iid Gaussian noise in the data domain to iid Gaussian noise in the measurement domain:

Lemma 1. *Let ε be a vector with $\mathcal{C}$ iid $\mathcal{N}(0, \sigma_d^2)$ entries and e be a vector with N iid $\mathcal{N}(0, \sigma_m^2)$ entries. Let $\hbar = \Phi\varepsilon$ and $\nu = \hbar + e$. Then ν contains N entries, sampled iid from $\mathcal{N}\left(0, \sigma^2\right)$, where $\sigma^2 = \frac{\mathcal{C}}{N}\sigma_d^2 + \sigma_m^2$.*

Proof: The tight frame property implies

$$\mathbb{E}\left[\hbar\hbar^\dagger\right] = E[\Phi\varepsilon\varepsilon^\dagger\Phi^\dagger] = \sigma_d^2\Phi\Phi^\dagger = \frac{\mathcal{C}}{N}\sigma_d^2\, I.$$

Therefore, $\nu = \hbar + e$ contains iid Gaussian elements with zero mean and variance σ^2. ■

Next we construct two families of low-coherence tight frames from Delsarte-Goethals codes.

C. Delsarte-Goethals Sets of Binary Symmetric Matrices

The finite field $\mathbb{F}_{2^m}$ is obtained from the binary field $\mathbb{F}_2$ by adjoining a root ξ of a primitive irreducible polynomial g of degree m. The elements of $\mathbb{F}_{2^m}$ are polynomials in ξ of degree at most $m-1$ with coefficients in $\mathbb{F}_2$, and we will identify the polynomial $x_0 + x_1\xi + \cdots + x_{m-1}\xi^{m-1}$ with the binary m-tuple $(x_0, \cdots, x_{m-1})$. The *Frobenius map* $f : \mathbb{F}_{2^m} \to \mathbb{F}_{2^m}$ is defined by $f(x) = x^2$ and the *Trace map* $\mathrm{Tr} : \mathbb{F}_{2^m} \to \mathbb{F}_2$ is defined by

$$\mathrm{Tr}(x) \doteq x + x^2 + \cdots + x^{2^{m-1}}.$$

The identity $(x+y)^2 = x^2 + y^2$ implies that $\mathrm{Tr}(x+y) = \mathrm{Tr}(x) + \mathrm{Tr}(y)$; the trace is a linear map over the binary field $\mathbb{F}_2$. The trace inner product given by $(v,w) = \mathrm{Tr}(vw)$ is non-degenerate; if $\mathrm{Tr}(vz) = 0$ for all z in $\mathbb{F}_2^m$ then $v = 0$. Every element a in $\mathbb{F}_{2^m}$ determines a symmetric bilinear form $\mathrm{Tr}[xya]$ to which is associated a binary symmetric matrix $P^0(a)$.

$$\mathrm{Tr}[xya] \doteq (x_0 \cdots x_{m-1}) P^0(a) (y_o \cdots y_{m-1})^\top.$$

The *Kerdock set* $\boldsymbol{K_m}$ is the m-dimensional binary vector space formed by the matrices $P^0(a)$. For example, let $m = 3$, and assume the finite field $\mathbb{F}_8$ is generated by adjoining a root ξ of the polynomial $g(x) = x^3 + x + 1$. Then K_3 is spanned by

$$P^0(100) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad P^0(010) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \text{ and } P^0(001) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Theorem 4. *Every nonzero matrix in K_m is nonsingular.*

Proof: If $xP^0(a) = 0$ then $\mathrm{Tr}[xya] = 0$ for all $y \in \mathbb{F}_{2^m}$. Now the non-degeneracy of the trace implies $a = 0$. ∎

Next we define higher order bilinear forms, each associated with a binary symmetric matrix. Given a positive integer t where $0 < t < \frac{m-1}{2}$ and given a field element a

$$\mathrm{Tr}\left[\left(xy^{2^t} + x^{2^t}y\right)a\right]$$

defines a symmetric bilinear form that is represented by a binary symmetric matrix $P^t(a)$ as above:

$$\mathrm{Tr}\left[\left(xy^{2^t} + x^{2^t}y\right)a\right] \doteq (x_0 \cdots x_{m-1}) P^t(a) (y_o \cdots y_{m-1})^\top \tag{3}$$

The *Delsarte-Goethals set* $DG(m,r)$ is then defined as

$$DG(m,r) \doteq \left\{ \sum_{t=0}^{r} P^t(a_t) \,|\, a_t \in \mathbb{F}_{2^m},\ t = 0, 1, \cdots, r \right\}.$$

The Delsarte-Goethals sets are nested

$$K_m = DG(m,0) \subset DG(m,1) \subset \cdots \subset DG\left(m, \frac{m-1}{2}\right),$$

and every bilinear form is associated with some matrix in $DG\left(m, \frac{m-1}{2}\right)$.

For example, let $m = 3$ and $g(x) = x^3 + x + 1$, the set $DG(3,1)$ is spanned by K_3, and

$$P^1(100) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad P^1(010) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } P^1(001) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Theorem 5. *Every nonzero matrix in $DG(m,r)$ has rank at least $m - 2r$.*

Proof: If x is in the null space of $\sum_{t=0}^{r} P^t(a_t)$, then for all $y \in \mathbb{F}_{2^m}$

$$\mathrm{Tr}\left[xya_0 + \sum_{t=1}^{r} \left(xy^{2^t} + x^{2^t}y\right) a_t\right] = 0.$$

Since $\mathrm{Tr}(x) = \mathrm{Tr}(x^2) = \cdots = \mathrm{Tr}\left(x^{\frac{1}{2}}\right)$ we have

$$\mathrm{Tr}\left[\left((xa_0)^{2^r} + \sum_{t=1}^{r} (xa_t)^{2^{t-r}} + a_t^{2^r}\, x^{2^{t+r}}\right) y^{2^r}\right] = 0.$$

Non-degeneracy of the trace now implies

$$(xa_0)^{2^r} + \sum_{t=1}^{r} (xa_t)^{2^{t-r}} + a_t^{2^r}\, x^{2^{t+r}} = 0.$$

This is a polynomial of degree at most 2^{2r} so there are at most 2^{2r} solutions. Hence the rank of the binary symmetric matrix $\sum_{t=0}^{r} P^t(a_t)$ is at least $m - 2r$. ■

III. Delsarte-Goethals Sensing

A. Delsarte-Goethals Frames

We start by picking an odd number m. The 2^m rows of the sensing matrix Φ are indexed by the binary m-tuples x, and the $2^{(r+2)m}$ columns are indexed by the pairs P, b, where P is an $m \times m$ binary symmetric matrix in the Delsarte-Goethals set $DG(m, r)$, and b is a binary m-tuple. The entry $\varphi_{P,b}(x)$ is given by

$$\varphi_{P,b}(x) = \frac{1}{\sqrt{N}} i^{xPx^\top + 2bx^\top} \tag{4}$$

Note that all arithmetic in the expressions $xPx^\top + 2bx^\top$ takes place in the ring of integers modulo 4. Given P, b the vector $xPx^\top + 2bx^\top$ is a codeword in the Delsarte-Goethals code (defined over the ring of integers modulo 4). For a fixed matrix P, the 2^m columns $\varphi_{P,b}\ ,\ b \in \mathbb{F}_2^m$ form an orthonormal basis. The name Delsarte-Goethals frame (DG frame) reflects the fact that Φ is a union of orthonormal bases. Hence, it is a tight-frame with redundancy $\frac{\mathcal{C}}{N}$. Delsarte-Goethals frames are highly incoherent (see [17]):

Proposition 2. *Let m and r be non-negative integers where m is odd and $r < \frac{m-1}{2}$. Then the worst case coherence μ of the sensing matrix derived from the $DG(m, r)$ set satisfies $\mu \leq \frac{1}{N^{\frac{1}{2} - \frac{r}{m}}}$.*

Sensing matrices derived from Delsarte-Goethals sets are incoherent tight frames so the results of Section §II can be brought to bear. The $N \times N^2$ sensing matrix derived from the Kerdock set is the union of N mutually unbiased bases and the worst case coherence matches the lower bound derived by Levenshtein [18] (see also Strohmer and Heath [8]).

B. Delsarte-Goethals Sieves

Chirp Detection [17] and Witness Averaging [19] are fast reconstruction algorithms that exploit the structure of Delsarte-Goethals frames. By sieving the testimony of witnesses [19] it is possible to detect the presence or absence of a signal at any given position in the data domain without explicitly reconstructing the entire signal.

There is however an aliasing problem with DG frames. When two signals modulate columns in the same orthonormal basis, spurious tones are generated by both the chirp detection and witness interrogation algorithms. This can be resolved by decimating the DG frame so that no two columns share the same binary symmetric matrix P. The simplest way to do this is to retain columns

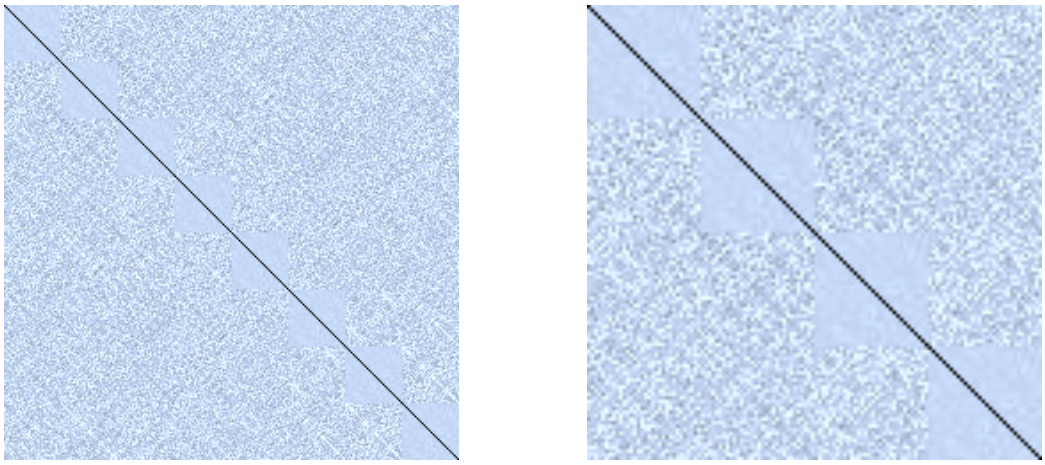
$$\varphi_P(x) = \frac{1}{\sqrt{N}} i^{xPx^\top}. \tag{5}$$

TABLE I: Spectral norms of $DG(m,1)$ frames and $DG(m,1)$ sieves as a function of m

$DG(m,1)$	$m=3$	$m=5$	$m=7$	$m=9$
Frame	2.8284	5.6569	11.3137	22.6274
Sieve	5.6568	11.1295	25.0386	55.0338

for which $b=0$. We call these subsampled matrices Delsarte-Goethals sieves ($DG(m,r)$ sieves) since it is still possible to sieve the testimony of witnesses. Note that each column of a DG sieve, is a column of the corresponding DG sieve, and the worst case coherence bound follows from Proposition 2. Figure 1 shows the distribution of the absolute value of pairwise inner products between columns of the $DG(5,1)$ sieve. All entries on the main diagonal are equal to 1, and around the the diagonal there are squares corresponding to translates of the Kerdock set K_m.

Table I shows that subsampling may increase the spectral norm. This will make it more difficult to reconstruct the signal either by chirp detection or by sieving the testimony of witnesses. We need to understand this increase in order to be able to apply the results of Section §II.



(a) Inner product between the first 512 columns of the $DG(5,1)$ matrix

(b) Inner product between the first 256 columns of the $DG(5,1)$ matrix

Fig. 1: The inner product between the columns of a $DG(5,1)$ matrix. The point at position (i,j) shows the inner product between the columns φ_i and φ_j. Lighter color shows higher inner product value.

C. Spectral Norm of DG Matrices

Given a sensing matrix, the results presented in Section §II show that if the the worst case coherence and spectral norm are sufficiently small then ℓ_0 minimization has a unique solution which coincides with the solution of a convex LASSO program. The worst case coherence μ of the initial $DG(m,r)$ frame satisfies $\mu \leq N^{\frac{r}{m}-\frac{1}{2}}$. To make sure that every row sum vanishes, we further exclude the $m+1$ rows, indexed by powers of 2, from the DG sieve. This exclusion changes the worst case coherence by at most

$\frac{m+1}{N}$ $\left(\text{Now } \mu \leq N^{\frac{r}{m}-\frac{1}{2}} + \frac{m+1}{N}\right)$. The experimental results presented below suggest that the number of pairs of rows in a DG sieve that fail to be orthogonal is very small. Removing these rows results in an equiangular tight frame that is not a union of orthonormal bases.

Table I lists the spectral norm of $DG(m,r)$ frames and $DG(m,r)$ sieves for $m = 3, 5, 7$ and 9. The spectral norm of a sieve is almost twice that of the corresponding frame and we shall see that the reason is a small number of duplicate rows. Removing these rows results in an equiangular tight frame. We now describe how to find these duplicate rows.

Let x, y be two distinct elements of the finite field $\mathbb{F}_2^m$, and let $\varphi(x)$, $\varphi(y)$ denote the two rows in Φ indexed by x and y. Setting $y = x + e$ we obtain

$$\begin{aligned} \varphi(x)^\dagger \, \varphi(y) &= \frac{1}{N} \sum_{P \in DG(m,r)} \imath^{(x+e)P(x+e)^\top - xPx^\top} = \frac{1}{N} \sum_{P \in DG(m,r)} \imath^{2ePx^\top + ePe^\top} \\ &= \frac{1}{N} \prod_{t=0}^{r} \left(\sum_{a \in \mathbb{F}_2^m} \imath^{2eP^t(a)xT\top + eP^t(a)eT\top} \right). \end{aligned} \tag{6}$$

If rows $\varphi(x)$ and $\varphi(y)$ are not orthogonal then each term in the product is nonzero. When $t > 0$ we now show that the t^{th} term in the product is a sum of linear characters. Since the index of summation ranges over the group, the sum is either zero or the linear character is trivial (each term in the sum is equal to 1).

Lemma 2. *Let $t \geq 1$ and let x and $x+e$ be two distinct elements of $\mathbb{F}_2^m$. Then either $\sum_{a \in \mathbb{F}_2^m} \imath^{e\,P^t(a)(2x+e)^\top}$ is zero, or for every field element a: $(x+e)P^t(a)(x+e)^\top - xP^t(a)x^\top = 0 \pmod 4$.*

Proof: When $t > 0$ every matrix $P^t(a)$ has zero diagonal and the map $a \to (e + 2x)P^t(a)e^\top$ is a linear map from the additive group $\mathbb{F}_2^m$ to $2\,\mathbb{Z}_4$. If this map is not identically zero then the character sum vanishes. ■

The next proposition follows from non-degeneracy of the trace.

Proposition 3. *If $t > 0$ then for every field element f*

$$f\,P^t(a)f^\top = 2\mathrm{Tr}\left(f^{2^t+1}\,a\right) + 2z_a f^\top \pmod 4 \quad \textit{where } z_a = \left[\mathrm{Tr}\left(\xi^{j(2^t+1)}\,a\right)\ j = 0, \cdots, m-1\right]. \tag{7}$$

Proof: Since the quadratic forms $fP^t(a)f^\top$ and $2\mathrm{Tr}\left(af^{2^t+1}\right)$ determine the same bilinear form they differ by a linear function $2z_a\,f^\top$. Since the quadratic form $fP^t(a)f^\top$ vanishes at all standard coordinate vectors we are able to determine the entries of the vector $2z_a$ that describes the linear function. ■

Next we use non-degeneracy of the trace to find duplicate rows $\varphi(x)$ and $\varphi(x+e)$.

Lemma 3. *The existence of field elements x, e such that*

$$(x+e)P^t(a)(x+e)^\top - xP^t(a)x^\top = 0 \,(\textit{mod } 4) \quad \textit{for all } a \textit{ in } \mathbb{F}_2^m, \tag{8}$$

is equivalent to the existence of a solution $\frac{x}{e}$ to the equation

$$1 + \frac{x}{e} + \left(\frac{x}{e}\right)^{2^t} + \sum_{j=0}^{m-1} e_j \left(\frac{\xi^j}{e}\right)^{2^t+1} = 0. \tag{9}$$

Proof: Since the trace is a linear map we may replace (8) by the condition that for all a in $\mathbb{F}_2^m$

$$\mathrm{Tr}\left[a\left((x+e)^{2^t+1}+x^{2^t+1}+\sum_{j=0}^{m-1}e_j\xi^{j(2^t+1)}\right)\right]=0.$$

Now the non-degeneracy of the trace implies that $(x+e)^{2^t+1}+x^{2^t+1}+\sum_{j=0}^{m-1}e_j\xi^{j(2^t+1)}=0$. Expanding $(x+e)^{2^t+1}$, we ontain

$$e^{2^t+1}+x\,e^{2^t}+x^{2^t}\,e+\sum_{j=0}^{m-1}e_j\xi^{j(2^t+1)}=0.$$

Since e is non-zero, dividing the equation by e^{2^t+1} completes the proof. ■

The solutions to the equation $z+z^{2^t}=0$ form a subfield of $\mathbb{F}_2^m$ and the number of solutions is $\gcd\left(2^t-1,2^m-1\right)$. Note that when m is odd and $t=1$ or $t=2$, there are exactly two solutions ($z=0$ and $z=1$). We now list the conditions satisfied by x and e if the row $\varphi(x)$ is not orthogonal to the row $\varphi(x+e)$.

Theorem 6. *Let x and $x+e$ be two distinct elements of the finite field $\mathbb{F}_2^m$. Then $\varphi(x)^\dagger\varphi(x+e)\neq 0$ if and only if the following conditions simultaneously hold:*

- *(C1) For every $t\geq 1$: $\frac{x}{e}+\left(\frac{x}{e}\right)^{2^t}=1+\sum_{j=0}^{m-1}e_j\left(\frac{\xi^j}{e}\right)^{2^t+1}$.*
- *(C2) $\sum_{a\in\mathbb{F}_2^m}\imath^{e\,P^0(a)(2x+e)^\top}\neq 0$.*

Theorem 6 provides an efficient way for identifying the non-orthogonal rows of the sieve matrices without requiring to calculate the gram matrices $\Phi^\dagger\Phi$ explicitly. For every element e, we first find the solution for the case $t=1$. If such a solution exists then we just need to *check* that condition (C1) is valid for other values of t. If all conditions passed then we just verify condition (C2). This method significantly reduces the computational cost of eliminating the non-orthogonal rows.

The next formula is for $t=1$

$$\frac{x}{e}+\left(\frac{x}{e}\right)^2=\lambda\quad\text{where }\lambda=1+\frac{\sum_{j=0}^{m-1}e_j\xi^{3j}}{e^3}.$$

This is a quadratic equation with roots $\frac{x}{e}$ and $\frac{x}{e}+1$ where $\frac{x}{e}\doteq\sum_{\substack{\ell:\,odd\\1\leq\ell\leq m-2}}\lambda^{2^\ell}$. On the other hand

$$\lambda+\lambda^2=\frac{x}{e}+\left(\frac{x}{e}\right)^4=\alpha\quad\text{where }\alpha=1+\frac{\sum_{j=0}^{m-1}e_j\xi^{5j}}{e^5}.$$

Thus we can also retrieve the explicit solution $\lambda=\sum_{\substack{\ell:\,odd\\1\leq\ell\leq m-2}}\alpha^{2^\ell}$. In other words, the following equivalence between the two field elements (which are both functions of e) must be satisfied:

$$\sum_{\substack{\ell:\,odd\\1\leq\ell\leq m-2}}\left(1+\frac{\sum_{j=0}^{m-1}e_j\xi^{5j}}{e^5}\right)^{2^\ell}=1+\frac{\sum_{j=0}^{m-1}e_j\xi^{3j}}{e^3}.\tag{10}$$

Remark 1. *Solutions to condition (C1) correspond to codewords of weight 2 in the binary code that is dual to the code determined by matrices in $DG(m,r)$ with zero diagonal. The number of solutions can be calculated using the MacWilliams Identities and we provide details in Appendix §A.*

Table II records the number of duplicate measurements that need to be deleted in order to transform a $DG(m,1)$ sieve into a tight frame. We calculated the number of duplicate rows for $DG(m,2)$, where

$m \leq 15$, and found that there were no solutions to (10) that also satisfied (C2); that is all $DG(m,2)$ sieves with $m \leq 15$ are tight frames. Hence

Conjecture: Every $DG(m,r)$ sieve with $r \geq 2$ is a tight-frame.

Figure 2 displays for $m = 7$ and 9 the average condition number of a random $N \times k$ submatrix of the $DG(m,1)$ sieve and the $DG(m,0)$ frame. The spectral norm of the hollow gram matrix $\|\Phi^\dagger\Phi - \mathrm{I}_N\|_2$ was calculated for 2000 randomly chosen submatrices Φ_k and the average was recorded. The comparison with Gaussian sensing matrices was made by drawing 10 iid Gaussian matrices, calculating for each matrix the average spectral norm over randomly chosen submatrices, and then recording the median value.

TABLE II: Number of row deletions required to transform a $DG(m,1)$ sieve into a tight frame.

$DG(m,1)$	$m=5$	$m=7$	$m=9$	$m=11$	$m=13$	$m=15$
# of non-orthogonal rows	11	25	45	83	203	381
% of non-orthogonal rows	0.3438	0.1953	0.0879	0.0405	0.0248	0.0116

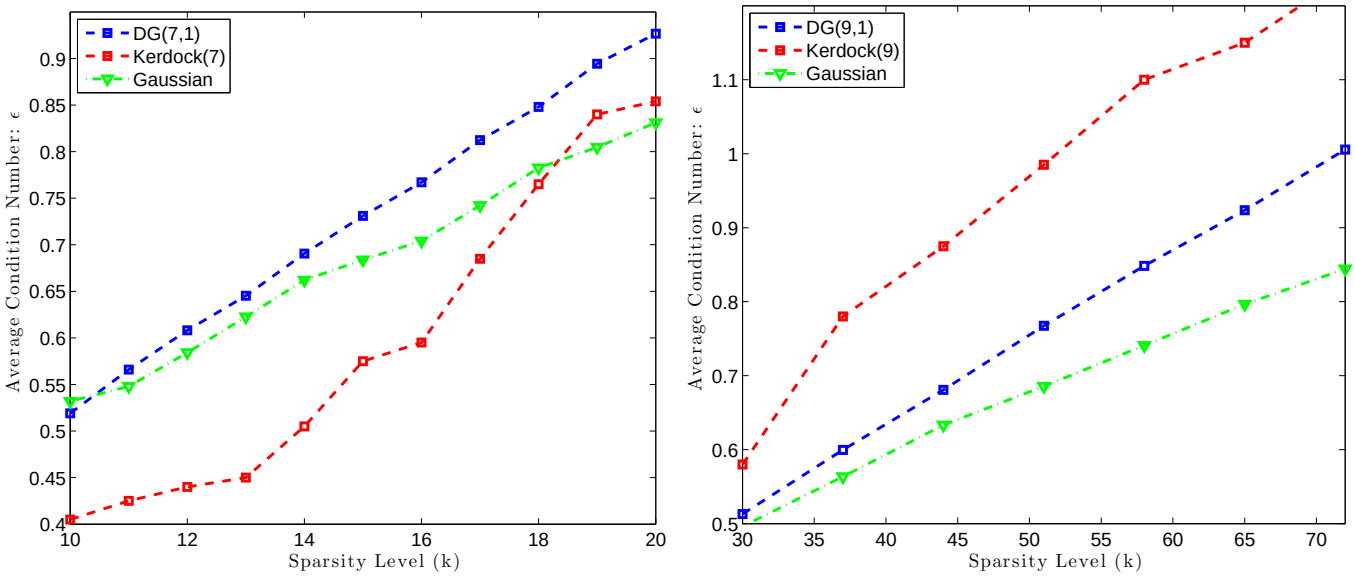


Fig. 2: Average spectral norm of $\Phi_k^\dagger\Phi_k - \mathrm{I}_{k\times k}$, where Φ_k is a random sub dictionary of Φ. Here the comparison is between Gaussian, $DG(m,1)$ sieve, and $DG(m,0)$ base matrices. Each experiment is repeated 2000 times.

Remark 2. Here we compare the empirical results of Figure 2 with the theoretical results of Theorem 2. First we considered the $DG(7,0)$ frame, with $\mathcal{C} = 2^{14}$ and $N = 2^7$. The worst case coherence of Φ is $\mu = 2^{-\frac{7}{2}}$, and the square of the spectral norm of Φ is 2^7. So the constant c in Theorem 3 needs to be at least $\mu \log \mathcal{C} = \frac{14\log 2}{8\sqrt{2}} \approx 0.85$. Hence, as long as k is at most $\frac{0.85\times 128}{14\log 2} \approx 11$, Theorem 2 predicts probability of non-uniqueness on the order of 2^{-14}. Experimental results presented in Figure 2a are more positive; all 2000 trials resulted in sub-dictionaries with full rank, even for k as large as 20.

Next we considered the $DG(7,1)$ sieve with $\mathcal{C} = 2^{14}$ and $N = 103$[1]. The worst case coherence of Φ is

[1]The 25 duplicate rows were removed from the matrix.

$\mu \approx 2^{-\frac{5}{2}}$, and the square of the spectral norm of Φ is $\|\Phi\|^2 \approx \frac{16384}{103} = 159.6$. As a result, the constant c needs to be at least $\frac{14 \log 2}{4\sqrt{2}} \approx 1.70$. Therefore, as long as k is less than $\frac{1.70 \times 103}{14 \log 2} \approx 10$ Theorem 2 predicts probability of non-uniqueness on the order of 2^{-14}. Again, we see that the theoretical bound is not tight, and for k as large as 20 all trials provide uniqueness of sparse representation.

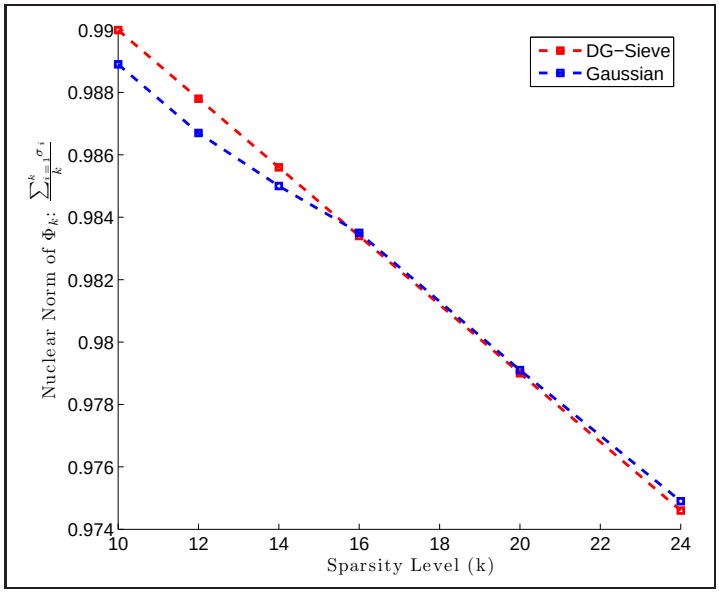


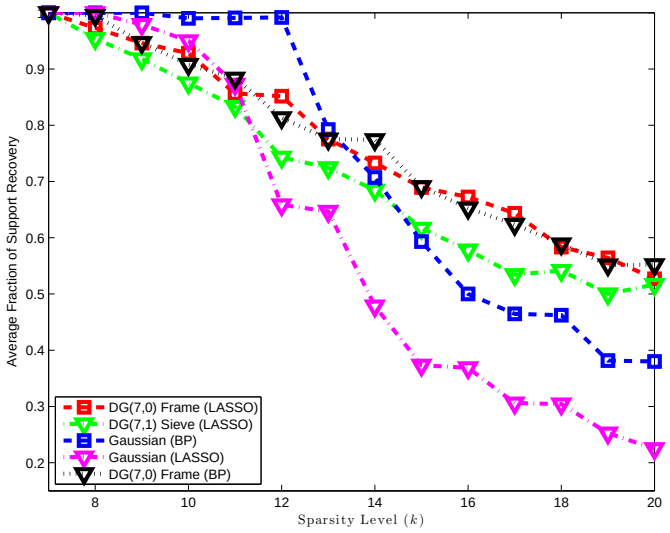
Fig. 3: Average nuclear norm $\left(\frac{1}{k}\sum_{i=1}^{k}\sigma_i\right)$ of random sub-dictionaries of of $DG(7,1)$ and Gaussian matrices of the same size as a function of the sparsity level k.

Remark 3. The bounds of Proposition 1 only apply to the condition number of random submatrices and do not provide additional information about the distribution of eigenvalues. However Gurevich and Hadani [20] have analyzed the spectrum of certain incoherent dictionaries that are unions of disjoint orthonormal bases. They have shown that the eigenvalues of the Gram matrix of a random subdictionary are asymptotically distributed around 1 according to the Wigner semicircle law. Our experimental results suggest that this property is shared by DG sieves which are not unions of orthonormal bases. Figure 3 shows that the distribution of the singular values of a random submatrix of a DG sieve is symmetric around 1, and very similar to the distribution for a Gaussian matrix of the same size.

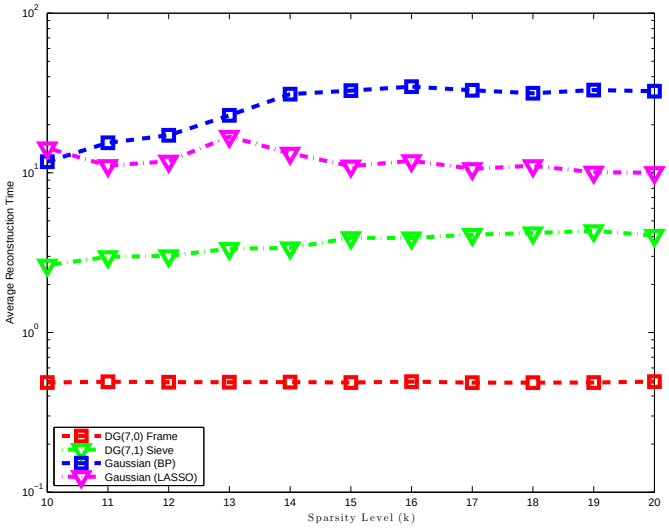
IV. Numerical Experiments

In this Section we present numerical experiments to evaluate the performance of the DG frames and sieves. The performance of DG frames and sieves is compared with that of random Gaussian sensing matrices of the same size. The SpaRSA algorithm [15] with ℓ_1 regularization parameter $\lambda = 10^{-9}$ is used for signal reconstruction in the noiseless case, and the parameter is adjusted according to Theorem 3 in the noisy case. The reason for using SpaRSA is that is designed to solve complex valued LASSO programs.

Remark 4. Given a random sensing matrix satisfying RIP, it is known that Basis Pursuit leads to more accurate reconstruction than the LASSO [1]. It is for this reason that we also compare results for LASSO applied to DG matrices with results for Basis Pursuit applied to Gaussian matrices. The ℓ_1-magic package[21] is used to solve the Basis Pursuit optimization program. The results for Gaussian matrices shown in Figure 4 are consistent with the observation made in [22] that when the signal is

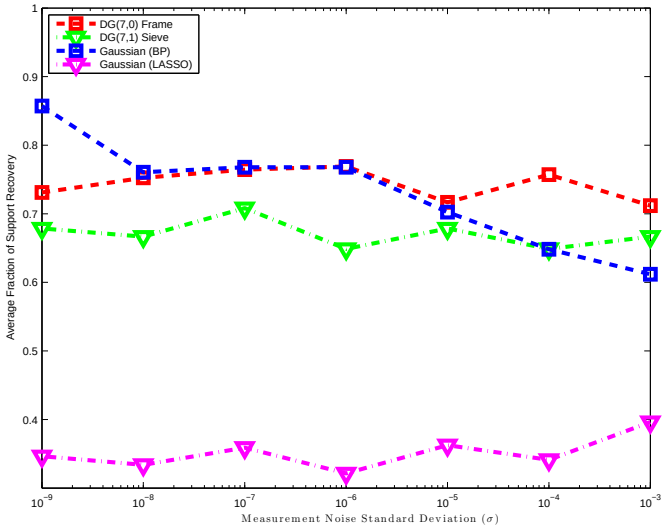


(a) Average fraction of the support that is reconstructed successfully as a function of the sparsity level k

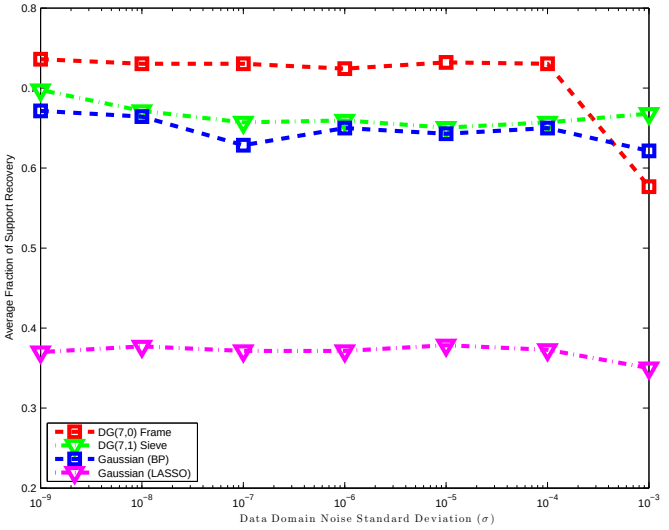


(b) Average reconstruction time in the noiseless regime for different sensing matrices.

Fig. 4: Comparison between $DG(7,0)$ frame, $DG(7,1)$ sieve, and Gaussian matrices of the same size in the noiseless regime. The regularization parameter for LASSO is set to 10^{-9}.



(a) The impact of the noise in the measurement domain on the accuracy of the sparse approximation for different sensing matrices.



(b) The impact of the noise in the data domain on the accuracy of the sparse approximation for different sensing matrices.

Fig. 5: Average fraction of the support that is reconstructed successfully as a function of the noise level in the measurement domain (left), and in the data domain (right). Here the sparsity level is 14. The regularization parameter for LASSO is determined as a function of the noise variance according to Theorem 3.

not very sparse, interior point methods (ℓ_1 - magic) are less sensitive than gradient descent methods (SpaRSA)

For Gaussian matrices, we sampled 10 iid random matrices independently to eliminate the exponentially small chance of getting a sample Φ with $\mu = \omega(N)$ or $\|\Phi\|^2 = \omega\left(\frac{\mathcal{C}}{N}\right)$, and the median of the results

among all 10 random matrices is reported. The use of 10 random trials to eliminate pathological sensing matrices is standard practice (see [11] for example).

The experiments relate accuracy of sparse recovery to the sparsity level and the Signal to Noise Ratio (SNR). Accuracy is measured in terms of the statistical $0-1$ loss metric which captures the fraction of signal support that is successfully recovered. The reconstruction algorithm outputs a k-sparse approximation $\hat{\alpha}$ to the k-sparse signal α, and the statistical $0-1$ loss is the fraction of the support of α that is not recovered in $\hat{\alpha}$. Each experiment was repeated 2000 times and Figure 4 records the average loss.

Figure 4 plots statistical $0-1$ loss and complexity (average reconstruction time) as a function of the sparsity level k. We select k-sparse signals with uniformly random support, with random signs, and with the amplitude of non-zero entries set equal to 1. Three different sensing matrices are compared; a Gaussian matrix, a $DG(7,0)$ frame and a $DG(7,1)$ sieve. After compressive sampling the signal support is recovered using the SpaRSA algorithm with $\lambda = 10^{-9}$. For random matrices the signal support is also recovered by ℓ_1-minimization.

Figure 5a plots statistical $0-1$ loss as a function of noise in the measurement domain and Figure 5b does the same for noise in the data domain. In the measurement noise study, a $\mathcal{N}(0,\sigma^2)$ iid measurement noise vector is added to the sensed vector to obtain the N dimensional vector f. The original k-sparse signal α is then approximated by solving the LASSO program with $\lambda = 2\sqrt{2\log\mathcal{C}}\sigma^2$, and basis pursuit with $\epsilon = 2N\sigma^2$. Following Lemma 1, we use a similar method to study noise in the data domain. Figure 5 shows that DG frames and sieves outperform random Gaussian matrices in terms of noisy signal recovery using the LASSO.

V. Conclusion

We have constructed two families of deterministic sensing matrices, $DG(m,r)$ frames and $DG(m,r)$ sieves, by exponentiating codewords from $\mathbb{Z}_4$ - linear Delsarte-Goethals codes. We have verified that the worst-case coherence and the spectral norm of these sensing matrices satisfy the conditions necessary for uniqueness of sparse representation and fidelity of ℓ_1 reconstruction via the LASSO algorithm. We have presented numerical results that confirm performance predicted by the theory. These results show that DG frames and sieves outperform random Gaussian matrices in terms of noiseless and noisy signal recovery using the LASSO. Our focus here is on ℓ_1 reconstruction using the LASSO algorithm but we note that the particular structure of the DG matrices leads to faster algorithms and to additional features such as local decoding and stronger guarantees on resilience to noise in the data domain.

Acknowledgements

The authors would like to thank Marco Duarte and Waheed Bajwa for sharing many valuable insights, and Waheed in particular for his help with the SpaRSA package.

References

[1] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics, Vol. 59 (8) , pp. 1207-1223,* 2006.

[2] D. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory, Vol. 52 (4), pp. 1289-1306,* April 2006.

[3] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Transactions on Information Theory, Vol. 52 (12), pp. 5406-5425,* December 2006.

[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory, Vol. 52 (2), pp. 489-509,* 2006.

[5] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of American Mathematical Society Vol. 22, pp. 211-231,* 2009.

[6] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation, Vol 28 (3), pp. 253-263,* December 2008.

[7] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *Journal of Complexity, Vol. 23 (4-6), pp. 918-925*, August-December 2007.
[8] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," *Applied and Computational Harmonic Analysis, Vol. 14 (3), pp. 257275s*, May 2003.
[9] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," *Statistical Signal Processing. IEEE/SP 14th Workshop on Publication, pp. 294-298*, August 2007.
[10] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient compressed Sensing using Optimized Expander Graphs," *IEEE Transactions on Information Theory, Vol. 55 (9), pp. 4299-4308.*, 2009.
[11] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery.," *46th Annual Allerton Conference on Communication, Control, and Computing, pp. 798-805*, September 2008.
[12] V. Chandar, "A negative result concerning explicit matrices with the restricted isometry property," *Preprint*, 2008.
[13] J. Tropp, "The Sparsity Gap: Uncertainty Principles Proportional to Dimension," *To appear, Proc. 44th Ann. IEEE Conf. Information Sciences and Systems (CISS)*, 2010.
[14] E. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Annals of Statistics, Vol. 37, pp. 2145-2177*, 2009.
[15] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing, Vol. 57 (7), pp. 2479-2493*, July 2009.
[16] R. Calderbank, S. Howard, and S. Jafarpour, "Sparse reconstruction via the Reed-Muller sieve," *accepted to the International Symposium on Information Theory (ISIT)*, 2010.
[17] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of Matrices satisfying a Statistical Isometry Propery," *IEEE Journal of Selected Topics in Signal Processing, Special Issues on Compressive Sensing, , Vol. 4 (2), pp. 358-374*, 2010.
[18] V.I. Levenshtein, "Bounds on the maximum cardinality of a code with bounded modulus of the inner product," *Soviet Math. Dokl. Vol. 25, pp.526-531*, 1982.
[19] R. Calderbank, S. Howard, and S. Jafarpour, "A sub-linear algorithm for Sparse Reconstruction with $\ell 2/\ell 2$ Recovery Guarantees," *Preprint*, 2009.
[20] Sh. Gurevich and R. Hadani, "The statistical restricted isometry property and the Wigner semicircle distribution of incoherent dictionaries," *submitted to the Annals of Applied Probability*, 2009.
[21] E. Candès and J. Romberg, "ℓ_1-magic: Recovery of sparse signals via convex programming," *available at http://www.acm.caltech.edu/l1magic*, 2005.
[22] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Technical Report No. 2009-01, California Institute of Technology*, 2009.
[23] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland: Amsterdam, 1977.

APPENDIX A
THE NUMBER OF SOLUTIONS OF CONDITION (C1)

Let $DG_0(m,r)$ denote the set of all zero-diagonal matrices in $DG(m,r)$:

$$DG_0(m,r) = \left\{ \sum_{t=1}^{r} P^t(a_t) \,|a_t \in \mathbb{F}_2^m \, t = 1, \cdots, r \right\}.$$

For every matrix P in $DG_0(m,r)$, the vector $xPx^\top$ is a codeword of the linear binary code $\overline{DG}_0(m,r)$ which is a sub-code of the Delsarte-Goethals code. Note that $\overline{DG}_0(m,r)$ has 2^{rm} codewords of length 2^m. The following lemma shows how the number of solutions to (C1) is related to the properties of this binary code.

Lemma 4. *Let $\{W_0, \cdots, W_N\}$ denote the weight distribution of $\overline{DG}_0(m,r)$. Then the number of pairs $(x, x+e)$ satisfying (C1) is equal to*

$$\frac{1}{2^{rm}} \sum_{i=0}^{N} W_i \mathcal{K}_2(i), \tag{11}$$

where $\mathcal{K}_\ell(z)$ is the ℓ^{th} Krawtchouk polynomial, defined as

$$\mathcal{K}_\ell(z) = \sum_{r=0}^{\ell} \binom{z}{r} \binom{N-z}{\ell - r} (-1)^r. \tag{12}$$

Proof: Lemma 3 implies that the number pairs $(x, x+e)$ satisfying Condition (C1) is equal to the number of duplicate rows in $\overline{DG}_0(m,r)$. The condition that the rows x and $x+e$ are identical is equivalent to the condition that the vector with entry 1 in positions x and $x+e$, and zero elsewhere belongs to the dual code. The lemma now follows from the MacWilliams Identities [23] that relate relate the number of codewords of weight 2 in the dual of $\overline{DG}_0(m,r)$ to the weight distribution of $\overline{DG}_0(m,r)$. ■

Next we show that for the case $r=1$, the number of solutions to (C1) only depends on the number of codewords with weight 2^{m-1} in $\overline{DG}_0(m,1)$:

Theorem 7. *Let m be an odd number and let r equal 1. Then the number of solutions to (C1) is 2^m-1-s where s is the number of codewords with weight 2^{m-1} in $\overline{DG}_0(m,1)$.*

Proof: We start by calculating the rank of matrices in $DG_0(m,1)$: Let a be a fixed element of $\mathbb{F}_2^m$. A field element x is in the null space of P_a if and only if for every field element y, $xP_ay^\top = 0$. Using Equation 3, this condition can be translated to the condition

$$\mathrm{Tr}\left((xy^2+x^2y)a\right) = 0 \text{ for all } y.$$

Since $\mathrm{Tr}(x) = \mathrm{Tr}(x^2)$ the condition further reduces to

$$\mathrm{Tr}\left((xa+x^4a^2)y^2\right) = 0 \text{ for all } y.$$

Non-degeneracy of the trace implies that $x^4 + \frac{x}{a} = 0$, which, since m is odd, has the unique solution $x^3 = \frac{1}{a}$.

Now let $S = \sum_{x\in\mathbb{F}_2^m} \imath^{xP_ax^\top}$. Since $xP_ax^\top$ is a binary codeword, we have $S^2 = (N-2w_a)^2$, where w_a is the weight of the codeword determined by P_a. It has been proved in [17] that $S^2 = 2^m \sum_{e:eP_a=0} \imath^{eP_ae^\top}$. We provide the proof here for completeness:

We have

$$S^2 = \sum_{x,y} \imath^{xP_ax^\top + yP_ay^\top} = \sum_{x,y} \imath^{(x+y)P_a(x+y)^\top + 2xP_ay^\top}$$

Changing variables to $z = x\oplus y$ and y gives

$$S^2 = \sum_z \imath^{zP_az^\top} \sum_y (-1)^{zP_ay^\top} = 2^m \sum_{z:zP_a=0} \imath^{zP_az^\top}.$$

The null space of P_a has only two elements 0 and $a^{-\frac{1}{3}}$. As a result

$$S^2 = 2^m \left(1 + \imath^{a^{-\frac{1}{3}}P_a a^{\frac{1}{3}\top}}\right).$$

There are two cases; S^2 is either 0 or 2^{m+1}.

Case 1: S is zero. This case provides one possible weight value: $w_a = 2^{m-1}$.
Case 2: $|S|^2 = 2^{m+1}$. Therefore $2^m - 2w_a = \pm 2^{\frac{m+1}{2}}$. This case provides two distinct weight values: $w_a = 2^{m-1} \pm 2^{\frac{m-1}{2}}$.

Hence $DG_0(m,1)$ has exactly four distinct weights $\langle 0, 2^{m-1}-2^{\frac{m-1}{2}}, 2^{m-1}, 2^{m-1}+2^{\frac{m-1}{2}}\rangle$. Let $\langle 1,t,s,t'\rangle$ denote the corresponding weight distribution. We can use the MacWilliams identities to find the values of t and t' as a function of s. First, note that the dual code has exactly one codeword of weight 0. Using MacWilliams identities with Krawtchouk polynomial $\mathcal{K}_0(z) = 1$, gives the equation $1+t+s+t' = \mathcal{C}$. Second, since all matrices in $DG_0(m,r)$ are zero-diagonal, for every field element a and for every index j in $\{0,\cdots,m\}$, $\xi^j P_a {\xi^j}^\top = 0$, the dual code has exactly $m+1$ codewords of weight 1. Again,

MacWilliams identities, with Krawtchouk polynomial $\mathcal{K}_1(z) = N - 2z$ gives the equation $(m+1)N = N + \sqrt{2N}(t' - t)$. This equation can be simplified to $t - t' = m\,2^{\frac{m-1}{2}}$. Solving t and t' with respect to s gives $t = \frac{2^m - 1 - s + m2^{\frac{m-1}{2}}}{2}$ and $t' = \frac{2^m - 1 - s - m2^{\frac{m-1}{2}}}{2}$. The theorem then follows from substituting the values t, s, t' into Equation (12), and simplifying the expression using the Krawtchouk polynomial $\mathcal{K}_2(z) = \frac{(N-2z)^2 - N}{2}$. ■